Investment Performance Evaluation

Wayne E. Ferson

Working Paper Series

Investment Performance Evaluation

Wayne E. Ferson

Abstract: This paper provides a review of the rapidly developing literature on investment performance evaluation. The goals are to summarize the significant forces and contributions that have brought this field of research to its current state of knowledge and to suggest directions for future research. This review is written for readers who are familiar with financial economics but not the specific literature and who wish to become familiar with the current state of the art. Suggestions for future research include refinements to portfolio holdings–based performance measures, a more balanced treatment of costs, and clientele-specific measures of investment performance.

Investment Performance Evaluation

1. INTRODUCTION

This is a good time for a review of the academic finance literature on evaluating investment performance. While the literature goes back to before the 1960s, with antecedents to the 1930s, recent years have witnessed explosive growth . It is now difficult even for a specialist to keep up, and anything approaching a complete survey would require ten times the page length allotted here. This is a selective review of the methods for measuring portfolio performance and evidence on professionally managed investment portfolios, emphasizing what I think are the most important findings that should influence future work in this area. The selection bias reflects the author's interests and admittedly limited knowledge of so vast an area. My up-front apology is offered to the specialist whose papers I have omitted.

Section 2 provides a little background. What explains the dramatic growth in the academic research on investment performance? I offer some conjectures. Section 3 selectively reviews the main performance measures, using the stochastic discount factor as a unifying theme. Section 4 provides an overview of issues in the literature. Section 5 provides more suggestions for future research, and Section 6 concludes.

2. THE RESEARCH ENVIRONMENT

Recent years have witnessed an explosion of new methods for and new evidence on investment performance evaluation. Several forces have contributed to this renaissance. One important factor is that the cost of producing investment performance research has declined. Early studies relied on proprietary or expensive commercial data bases for their fund performance figures, or researchers collected data by hand from published paper volumes. In 1997 the Center for Research in Security Prices introduced the CRSP mutual fund data base. The fund tracking firm Morningstar started making data available to academics at reasonable cost at about this time. Starting in about 1994, several data bases on hedge funds became

available to academic researchers. The Securities and Exchange Commission (SEC), through its Edgar system, has made detailed data on the holdings of institutional investors available to researchers at low cost. International data on institutional investors' portfolio holdings are recently available to academics through data vendors such as Abel Noser (e.g. Irvine et al., 2004). Scholars increasingly post their data on web pages and in journals' data appendices.

Regulatory changes have also resulted in more data for academics to study. For example, the SEC has required registered investment management companies to report their investment holdings semi-annually since 1985 and quarterly since 2004. Some funds voluntarily report their holdings monthly to Morningstar, starting in about 1992. The SEC has required mutual funds to report a specific benchmark for comparison with their returns since 1999. There is likely to be future regulation that generates additional interesting data for researchers. Of course, during the same period the costs of computing have also declined dramatically.

Another factor in the performance evaluation renaissance is demand. The demand for research on managed portfolio performance increased as mutual funds and related investment vehicles became more important to investors in the 1980's and 1990s. During this period, equity investment became widely popular, as 401(k) and other defined-contribution investment plans began to dominate defined-benefit plans in the United States. Under such plans, individuals make their own investment choices from a menu of employer-specified options. At the same time, "baby-boomers" reached an age where they had more money to invest, and new investment opportunities were developing for investors in Europe and Asia that increased the demand for professionally managed portfolio products. This period also witnessed an explosive growth in alternative investments, such as hedge funds and private equity vehicles. Innovation in investment research combined with innovation in investment practice, with mutually reinforcing feedback. Bill Sharpe's early work on capital asset pricing (CAPM, 1964) stimulated the first indexed fund products, offered by Wells Fargo in 1971 to institutions and by Jack Bogle in 1976

to retail investors. More recently, industry innovations such as fund families, exchange-traded funds, sector and life-cycle funds have stimulated academic research. See Tkac (2008) for a survey of recent mutual fund industry innovations.

In addition to the effects of lower costs and increased demand, the research community capable of conducting and publishing academic work on investment performance has grown. You, the reader, may be part of this growing community.

3. PERFORMANCE MEASURES

There are a large number of performance measures, but not enough space to review them all here. I describe a subset of the measures needed to illustrate the main issues. For more comprehensive reviews, see Aragon and Ferson (2006), Carino, Christopherson and Ferson (2009) and Goetzmann, Ingersoll, Spiegel and Welch (2007).

3.1 ALPHA

The most famous performance measure is Alpha. Although its antecedents go back at least to Coles (1933), it was Jensen's (1968, 1972) work based on the Capital Asset Pricing Model (Sharpe, 1964) that helped make alpha famous. Alpha is now so ubiquitous that it has become a generic, like Xerox$^{\text{TM}}$ or Google. Investment practitioners routinely discuss their strategies in terms of their quest for alpha, and the number of investment firms with alpha in their names is truly staggering.

A natural definition of alpha is based on the now common stochastic discount factor (SDF) approach. The SDF approach appeared as early as Beja (1971), but became the common language of empirical asset pricing during the 1980s. A *stochastic discount factor*, m_{t+1}, is a scalar random variable, such that the following equation holds:

$$E(m_{t+1} R_{t+1} - \underline{1} | Z_t) = 0, \tag{1}$$

where Rt+1 is the vector of primitive asset gross returns (payoff at time t+1 divided by price at time t), $\underline{1}$ is an N-vector of ones and Zt denotes the investment client's or public information at time t. The elements of the vector mt+1 Rt+1 may be viewed as "risk adjusted" gross returns. The returns are risk adjusted by "discounting" them, or multiplying by the stochastic discount factor, mt+1, so that the expected "present value" per dollar invested is equal to one dollar. We say that an SDF "prices" the primitive assets if equation (1) is satisfied. If Xt+1 is the payoff and Pt is the price, then Rt+1 = Xt+1/Pt and Equation (1) says that Pt = E{mt+1Xt+1|Zt}. In the language of Fama (1970), this says that the price fully reflects Zt.

An investment manager forms a portfolio of the primitive assets with gross return $R_{pt+1} = x(\Omega_t)' R_{t+1}$, where $x(\Omega_t)$ is the vector of portfolio weights and Ω_t is the manager's information at time t. If Ω_t is more informative than Z_t, the portfolio R_{pt+1} may not be priced through equation (1). That is, the manager may record "abnormal performance," or nonzero alpha. (In what follows I will drop the time subscripts unless they are needed for clarity.)

Gosten and Jagannathan (1994) and Chen and Knez (1996) were the first to develop SDF alphas for fund performance. For a given SDF we may define the SDF alpha for a fund with return R_p as:

$$\alpha_p \equiv E(m R_p | Z) - 1. \tag{2}$$

Theoretically the SDF is not unique unless markets are complete. Thus, the SDF alpha reflects the underlying model, and we can use (2) to summarize the major models and performance

measures.   To see how a manager with superior information can generate alpha, substitute $R_p = x(\Omega)'R$ into (2) and use the definition of covariance and Equation (1) to express:

$$\alpha_p = \text{Cov}(mR' ; x(\Omega)|Z). \tag{3}$$

Equation (3) says that alpha is the covariance of the portfolio manager's weights with the risk-adjusted returns of the assets, conditional on the client's information, summed across assets. Obviously, if the weights $x(\Omega)$ use only public information Z the alpha is zero.

To interpret alpha from the client's perspective, consider an example where m_{t+1} is the intertemporal marginal rate of substitution in consumption: $m_{t+1} = \beta\, u'(C_{t+1})/u'(C_t)$, where $u'(C)$ is the marginal utility of consumption. In this case, Equation (1) is the Euler equation which must be satisfied in equilibrium. If the consumer has access to a fund for which the conditional alpha is not zero he or she will wish to adjust the portfolio *at the margin*, purchasing more of the fund if alpha is positive and less if alpha is negative.

While the preceding argument can motivate α_p as a guide for clients' marginal investment changes, a positive alpha does not imply that a client, confronted with a new investment opportunity, would wish to purchase a discrete amount. There are examples to the contrary in the empirical literature. Ferson (2009) uses a time-additive utility function in a multiperiod model where the indirect value function is J(W,s) and the SDF is $m_{t+1} = \beta\, J_w(W_{t+1},s_{t+1})/u_c(C_t)$, where subscripts denote derivatives and s is a vector of state variables.   The model presents the client, who was initially at an optimum given the menu of N basic assets and information Z, with a new investment opportunity described as a managed portfolio with return $R_p = x(\Omega)'R$. The client adjusts by forming new consumption and portfolio choices, until the alpha is zero at the new optimum.   In this model the optimal discrete amount purchased is proportional to the SDF alpha of the fund (Proposition I).

The SDF alpha is also the risk adjusted excess return on the fund, relative to that of a benchmark portfolio that is assumed to be correctly priced by the SDF. If R_{Bt+1} is the benchmark, then Equation (2) implies $\alpha_p \equiv E(m [R_p - R_B]|Z)$. In empirical practice, alpha is almost always measured as the expected return of the fund in excess of a benchmark, $E(R_p - R_B|Z)$. For example, in the CAPM R_B is a fund-specific weighted average of the market portfolio and a risk-free asset. From the definition of alpha we have:

$$\alpha_p = E[m(R_p - R_B)|Z] = E(m|Z) \, E(R_p - R_B|Z) + Cov(m, R_p - R_B|Z). \tag{4}$$

Since $E(m|Z)$ is the inverse of the gross risk-free rate, we see that the expected return in excess of a benchmark is equivalent to the SDF alpha, if and only if $Cov(m, R_p|Z) = Cov(m, R_B|Z)$. This condition defines R_B as an "*Appropriate Benchmark*," a concept that will prove useful below.

Each definition of m brings along a version of alpha and corresponding appropriate benchmarks. To illustrate, Jensen's (1968, 1972) alpha follows from the CAPM. In that model, m_{t+1} is linear in the market portfolio return (Dybvig and Ingersoll, 1982), so an appropriate benchmarks is a combination of the market portfolio and a risk-free asset with gross return R_f, having the same market "beta," β_p, as the fund. This combination is $R_B = \beta_p R_m + (1-\beta_p)R_f$. The benchmark has the same covariance with m_{t+1} as the fund.[1] Jensen's alpha is the expected excess return of the fund over this benchmark: $\alpha_p = E[R_p - \{\beta_p R_m + (1-\beta_p) R_f\}] = E[R_p - R_f - \{\beta_p (R_m - R_f)\}]$. This also shows that Jensen's alpha is the intercept in a time-series regression of $R_p - R_f$ on $R_m - R_f$.

Asset pricing in the 1970's began to explore models in which exposure to more than a single market risk factor determines expected returns. Merton (1973), Long (1974) and the Arbitrage Pricing Theory (APT) of Ross (1976) are the classical examples. It follows from a

---

[1] That is, $m = a + bR_m$, so $Cov(R_m, m) = bVar(R_m)$ and $Cov(R_p, m) = bCov(R_p, R_m) = Cov(\beta_p R_m, m) = Cov(\{ \beta_p R_m + (1-\beta_p) R_f\}, m)$.

"multi-beta" model that m_{t+1} is linear in the vector of risk factors (e.g. Ferson, 1995). By the argument in footnote 1, a benchmark portfolio that has the same regression betas on the risk factors as the fund is an appropriate benchmark, because it has the same covariance with m_{t+1} as the fund.

The papers of Long and Merton suggested interest rates and inflation as risk factors, but their models did not fully specify the risk factors, and the APT specifies the factors only in a loose statistical sense. This leaves it up to empirical research to identify the risk factors. Chen, Roll and Ross (1986) empirically evaluated several likely economic factors, and Chen, Copeland and Mayers (1987) used these in an evaluation of equity mutual funds. Connor and Korajczyk (1988) showed how to extract statistical factors from stock returns in a fashion theoretically consistent with the APT, and Connor and Korajczyk (1986) and Lehmann and Modest (1987) used statistical factors in models for mutual fund performance evaluation.

Current investment management practice typically assumes that the benchmark portfolio is defined by the fund manager's investment "style." Roughly, style refers to a subset of the investment universe in which a manager is constrained to operate, such as small capitalization stocks versus large stocks, or "value" versus "growth" firms. This leads to the idea of "style exposures," similar to the risk exposures or betas implied by the multiple-beta asset pricing models. In this approach the Appropriate Benchmark has the same style exposures as the portfolio to be evaluated, and this is assumed to imply the same covariances with m_{t+1}. The style-based approach is reflected prominently in academic studies following Fama and French (1996) such as Carhart (1997).

Daniel, Grinblatt, Titman and Wermers (1997) further refine style-based alphas, measuring the characteristics of the individual stocks held by the fund. The characteristics include the market capitalization or size, a measure of value (the ratio of book value to market value), and the return over the previous year. For a given fund, the benchmark is formed by

matching the characteristics of the stocks in the portfolio held by the fund with "passive" portfolios constructed to have the same characteristics as the stocks. The assumption is that matching the characteristics implies matching the covariance with m_{t+1}. In some cases the style of a fund is captured using the returns of other managed portfolios in the same market sector. The benchmark portfolio is then a combination of a manager's peers' portfolios. Such benchmarks are routinely reported in the popular press for hedge funds, for example. With this approach the measured performance is a zero-sum game, as the average performance measured in the peer group must be zero. This approach can make it easier to control for costs and risks, to the extent that the portfolio and its peers are similar in these dimensions.

3.2 EVIDENCE ABOUT ALPHAS

There are many hundreds of papers providing evidence about alphas, but several broad themes have emerged. Alphas are empirically sensitive to the choice of the benchmark (Carlson (1970), Lehmann and Modest (1987), Fama and French, 2009) but estimates of alpha using different benchmarks typically have high cross-sectional correlation. Averaged across funds, alphas are typically negative, and often by an amount similar to the average fund's expense ratio (about 1% per year for US equity funds). The distribution of alphas across funds is mildly skewed to the left (Jensen (1968), Wermers, 2009). Alphas display little persistence over time, especially once you control for factors like momentum in the stocks held that generate persistence (Christopherson and Turner (1991), Carhart, 1997). The persistence in alpha is stronger among the poorly-performing funds. The fractions of funds with positive alpha estimates may have declined since the 1990s. However, it is not clear and still subject to debate, whether or not the extreme positive alphas estimated in a group of funds are significantly greater than zero when you properly account for the multiple comparisons (e.g. Kowsowski et al. (2006), Barras et al. (2009), Fama and French, 2009).

3.3 CONDITIONAL PERFORMANCE EVALUATION (CPE)

Traditional, or unconditional alphas compare returns and risks measured as averages over an evaluation period, and these averages are taken "unconditionally," or without regard to variations in the state of financial markets or the broader economy (formally, Z_t in (2) is taken to be a constant). In the conditional performance evaluation (CPE) approach, the state of the economy is measured using predetermined, public information variables. This takes the view that a managed portfolio strategy that can be replicated using readily available public information should not be judged as having superior performance, consistent with a version of semi-strong-form market efficiency as described by Fama (1970). However, by choosing the lagged variables Z_t, it is possible to set the hurdle for superior ability at any desired level of information.

The empirical model proposed by Ferson and Schadt (1996) is:

$$r_{p,t+1} = \alpha_p + \beta_o\, r_{m,t+1} + \beta'\, [r_{m,t+1} \otimes Z_t] + u_{pt+1} \tag{5}$$

where $r_{p,t+1}$ is the return of the fund in excess of a short term "cash" instrument, and Z_t is the vector of lagged conditioning variables, in demeaned form. The symbol $\otimes$ denotes the kronecker product, or element-by-element multiplication when $r_{m,t+1}$ is a single market index. A special case of Equation (5) is Jensen's alpha, where the terms involving Z_t are omitted.

Equation (5) assumes a time-varying fund beta, $\beta(Z_t)$, that may be modeled as a linear function of the public information Z_t: $\beta(Z_t) = \beta_o + \beta'Z_t$. The interaction terms $\beta'[r_{m,t+1} \otimes Z_t]$ control for common movements in the fund's conditional beta and the conditional expected benchmark return. The Appropriate Benchmark portfolio in this setting is the "naive" dynamic strategy, formed using the public information Z_t, that has the same time-varying conditional beta

as the portfolio to be evaluated. This strategy has a weight at time t on the market index equal to $\beta_o + \beta' Z_t$, and $\{1 - \beta_o - \beta' Z_t\}$ is the weight in safe asset or cash. Using the same logic as before, Equation (4) implies that α_p in the Ferson and Schadt model is the difference between the expected return of the fund and that of the "dynamic strategy" with the same conditional beta.

Christopherson et al (1998 a,b) propose a refinement of (5) to allow for a time-varying conditional alpha as well as a time-varying beta. This refinement of the model may have more power to detect abnormal performance if performance varies with the state of the economy.

When conditioning information is involved, SDF alphas differ from beta pricing model alphas. The conditional SDF alpha given Z is $\alpha(Z) = E(mR-1|Z)$ and the unconditional alpha is $\alpha_u = E(mR-1)$, so $E(\alpha(Z))=\alpha_u$. The conditional alpha of a beta pricing model, in contrast, is the SDF alpha divided by the gross risk-free rate. When the risk-free rate is time varying, Jensen's inequality implies that the expected value of the conditional alpha in the beta pricing model is not the unconditional alpha. Ferson and Schadt (1996) find that average conditional alphas and unconditional alphas from beta pricing models differ empirically for equity style funds.

3.4 CPE EVIDENCE

The literature on conditional performance evaluation has generated several key results. Ferson and Schadt (1996) find that funds' risk exposures change significantly in response to public information variables such as the levels of interest rates and dividend yields. Using conditional beta models Ferson and Schadt (1996) and Kryzanowski, Lalancette and To (1997) find that the distribution of mutual fund alphas is shifted to the right, relative to the unconditional alphas, and is centered near zero. Thus, conditional models tend to paint a more optimistic picture of mutual fund performance than unconditional models. This general pattern is confirmed by Zheng (1999), Ferson and Qian (2004) and others.

Ferson and Warther (1996) attribute much of the difference between unconditional and

conditional alphas to predictable flows of public money into funds. Higher inflows occurs when the public expects high returns, which leads to larger cash holdings at such times. Holding more cash when returns are predictably higher leads to lower unconditional performance, but does not affect the CPE measures. In pension funds, which are not subject to high frequency flows of public money, no overall shift in the distribution of fund alphas is found when moving to conditional beta models (Christopherson, et. al., 1998).

3.5 MARKET TIMING MODELS

The term "market timing" has two distinct meanings in the literature on investment performance. The classical use of the term refers to the ability of an investment manager to take on more exposure to the market before it goes up and to pull out of the market before it goes down. A second use of the term emerged in association with mutual fund scandals in the early 2000's. Mutual funds allow investors to trade their shares at the end of each day at a fixed net asset value (NAV) per share, determined on the basis of closing prices of the securities at the end of each day. If the NAV is partly based on stale prices, it will adjust to price changes with a lag that may be predictable. This affords traders an opportunity to trade in and out of the fund profitably, at the expense of the other shareholders . Such traders are said to be "market timers" in the fund's shares. In some cases fund management encouraged market timers, and even allowed software optimizing the timers' trades. A particularly egregious form," late trading," occurs if the fund illegally allows preferred traders to place their orders after the close of the markets and the determination of the NAV (e.g. Zitzewitz, 2006). See Qian (2009b) for a review and analysis of the late trading related fund scandals, which ultimately led the SEC and the New York State Attorney General to initiate litigation on more than 1,000 mutual funds.

Successful timing implies higher market betas when the market subsequently goes up, lower betas when it goes down, and thus a convex relation between the fund's return and the

market portfolio return. Classical, returns-based models of market-timing use convexity, and there are two main approaches to capturing the convexity.

In the model of Merton and Henriksson (1981), the convexity is modeled with put or call options. The Merton-Henriksson market timing regression is:

$$r_{pt+1} = a_p + b_p r_{mt+1} + \Lambda_p Max(r_{mt+1},0) + u_{t+1}. \qquad (6)$$

The coefficient Λ_p measures the market timing ability. If $\Lambda_p = 0$, the regression reduces to the market model regression used to measure Jensen's alpha. Thus, under the null hypothesis that there is no timing ability, the intercept is Jensen's alpha from the CAPM.

The Treynor-Mazuy (1966) market-timing model is a quadratic regression:

$$r_{pt+1} = a_p + b_p r_{mt+1} + \Lambda_p r_{mt+1}^2 + v_{t+1}. \qquad (7)$$

Admati, Bhattacharya, Pfleiderer, and Ross (1986) formalize the model, showing how it can be derived from a timer's optimal portfolio weight, assuming normal distributions and managers with exponential utility functions. They show that the timing coefficient Λ_p is proportional to the product of the manager's risk tolerance and the precision of the manager's signal about the future market returns. Thus, an aggressive manager with lower ability can generate the same timing coefficient as a better informed but more conservative manager. This motivates trying to separate the two effects. Admati, Bhattacharya, Pfleiderer, and Ross (1986) show how to separate risk aversion and signal quality by estimating regression (7) together with a regression for the squared residuals of (7), on the market excess return. Coggin, Fabozzi and Rahman (1993) implement this approach on equity mutual fund data.

The empirical evidence using the classical market timing models draws several

conclusions. Estimates of the market timing coefficients are typically either insignificantly different from zero, or a significant fraction of the coefficients are negative (Kon (1983), Henriksson and Merton (1984), Chang and Lewellen (1984), Becker, et al., 1999). A negative correlation in the cross-section between the intercepts and timing coefficients appears in several of the classical studies.

Early studies conclude that efforts at market timing by fund managers is likely wasted effort. However, the conclusions of the early studies seem problematical both on economic and on methodological grounds. On economic grounds, how do we understand the existence of a large number of asset allocation style mutual funds and pension funds using market timing strategies? From a methodological perspective, there are several reasons that we could observe convexity (or concavity) in the relation between the return of a fund and a benchmark that have nothing to do with market timing ability, as summarized by Chen, Ferson and Peters (2009). These include (1) underlying assets with nonlinearities (e.g., option holdings), (2) trading at higher frequency than returns are measured, or "interim trading," (3) fund trading in response to public information, and (4) systematically stale prices in funds' net asset values. Chen, Ferson and Peters (2009) incorporate controls for these other sources of nonlinearity in returns-based market timing models and find that they matter for measuring the timing ability of bond funds. In particular, the perverse negative timing coefficients are largely removed.

3.6 CONDITIONAL MARKET TIMING

Ferson and Schadt propose a CPE version of the Treynor Mazuy model:

$$r_{pt+1} = a_p + b_p r_{mt+1} + C_p'(Z_t r_{m,t+1}) + \Lambda_p r_{mt+1}^2 + w_{t+1}. \qquad (8)$$

In Equation (8), the term $C_p'(Z_t r_{m,t+1})$ controls for predictable time-variation in the market risk

premium and the fund's beta, just like it did in the regression (5). A manager who only uses Z_t has no conditional timing ability, and thus $\Lambda_p=0$. The coefficient Λ_p measures the market timing ability based on information beyond that contained in Z_t. Ferson and Schadt (1996) also develop a conditional version of the Merton and Henriksson (1984) market timing model. Further refinements are developed by Becker et al. (1999), Ferson and Qian (2004) and Chen, Ferson and Peters (2009).

While studies of mutual funds' market timing ability using the classical models found evidence of perverse, negative timing coefficients, Ferson and Schadt (1996) showed that the classical measures can produce negative coefficients for naïve dynamic strategies. The conditional timing measures can avoid this bias. Once standard public information variables are controlled for, there is little evidence that groups of mutual funds have conditional market-timing ability. However, in subsamples of asset-allocation style funds is there a hint of timing ability (Becker et al., 1999) and some evidence for conditional timing ability in certain economic conditions (Ferson and Qian, 2004).

Busse (1999) asks whether fund returns contain information about market volatility and develops volatility timing models. He finds evidence using daily data that funds may shift their market exposures in response to changes in second moments.

3.7 WEIGHT-BASED PERFORMANCE MEASURES

When the data on the holdings of the fund to be evaluated are available, it is possible to apply weight-based performance measures. These measures examine the relation between the manager's actual holdings and the subsequent returns of the assets. The idea is that a manager who increases the fund's portfolio weight in a security or asset class before it performs well, or who anticipates and avoids losers, has investment ability.

Weight-based performance measures go back at least to Cornell (1979) and Copeland and

Mayers (1982), Grinblatt and Titman (1993) develop the first measure for mutual funds. Applications to mutual funds include Grinblatt and Titman (1989a), Grinblatt, Titman, and Wermers (1995), Zheng (1999), Daniel, Grinblatt, Titman, and Wermers (1997), Kacperczyk, Sialm and Zheng (2005, 2008) and others. Applications to hedge funds include Griffin and Xu (2009) and Aragon and Martin (2009).

Grinblatt and Titman (1993) derive their measure in a single-period model where the fund managers maximizes the expected utility of the terminal wealth generated by the portfolio return, conditional on the information, Ω. When returns are conditionally normal given Ω, and assuming nonincreasing absolute risk aversion, they show that:

$$\text{Cov}\{ x(\Omega)'r \} > 0, \tag{9}$$

where $x(\Omega)$ is the optimal weight vector. Equation (9) says that the sum of the covariances between the weights of a manager with private information, Ω, and the returns for the securities in a portfolio is positive. If the manager has no information the covariance is zero.

From the definition of covariance we can implement (9) by demeaning the weights or the returns: $\text{Cov}\{ x(\Omega)'r \} = E\{[x(\Omega)-E(x(\Omega))] \ r\} = E\{x(\Omega) \ [r-E(r)]\}$. Copeland and Mayers (1982) and Ferson and Khang (2002) demean returns, while Brinson, Hood and Bebower (1986) and Grinblatt and Titman (1993) demean the weights, introducing a set of benchmark weights, x_B, as:

$$\text{Cov}\{x(\Omega)'r\} = E\{[x(\Omega)-x_B]'r \}. \tag{10}$$

With the benchmark weights, the benchmark portfolio implied by the weight based measure is given by $r_B= x_B'r$. Grinblatt and Titman (1993) define the benchmark as the fund's weights in the previous quarter. Thus, the model assumes that a manager with no information holds fixed

portfolio weights. Ferson and Khang (2002) define the benchmark weights as the portfolios' actual weights lagged k periods, updating these with a buy-and-hold strategy. Thus, each manager's position, k quarters ago, defines his "personal" benchmark. A manager with investment ability changes the portfolio in order to beat a buy-and-hold strategy. Fund performance is measured, according to Equation (10), as the average difference in raw returns over the subsequent quarter, between the fund and the benchmark portfolio defined by the weights, x_B. The return of the fund is a "hypothetical" return, since it is constructed using a snapshot of the fund's actual weights at the end of a period (usually, at the end of a quarter or half-year). This hypothetical return reflects no trading within the quarter, no trading costs or management fees.

Daniel, Grinblatt, Titman, and Wermers (DGTW, 1997) define the benchmark portfolio based upon the characteristics of the securities held. Specifically, each security in the fund's portfolio is assigned to one of 125 characteristic groups, depending upon its size, book-to-market ratio, and lagged return, measurable with respect to the beginning of the quarter. They construct passive (value-weighted) portfolios across all NYSE, AMEX, and NASDAQ stocks for each characteristic group. The return on the benchmark portfolio in a given quarter is the summation, across all securities in fund's portfolio, of the fund's portfolio weights times the return on the characteristics-matched portfolio for that stock. The DGTW alpha is the average difference between a fund's return and that of the DGTW benchmark.

Like the classical returns-based performance measures, unconditional weight-based measures have problems handling return dynamics. It is known that unconditional weight-based measures can show performance when the manager targets stocks whose expected return and risk have risen temporarily (e.g., stocks subject to takeover or bankruptcy); when a manager exploits serial correlation in stock returns or return seasonalities; and when a manager gradually changes the risk of the portfolio over time, as in style drift (Grinblatt and Titman, 1993). These problems

may be addressed using a conditional approach.

Ferson and Khang (2002) develop the *Conditional Weight-based Measure* of performance (CWM). Like other CPE approaches, the measure controls for changes in expected returns and volatility, as captured by a set of lagged economic variables or instruments, Z_t. The CWM uses the information in both the lagged economic variables and the fund's portfolio weights:

$$CWM = E\{ x(Z,\Omega)'[r - E(r|Z)] \}. \qquad (11)$$

The symbol $x(Z,\Omega)$ denotes the portfolio weight vector at the beginning of the period. The weights may depend on the public information, denoted by Z. The weights of a manager with superior information may also depend on the information, Ω. Here, we define the abnormal return as the component of return not expected by an investor who only sees the public information Z at the beginning of the period.

3.8 COMBINING HOLDINGS AND RETURNS

Grinblatt and Titman (1989a) study the reported holdings of mutual funds, in combination with the reported returns of the funds. Subsequent research has explored a number of decompositions that combine holdings with returns. The ingredients for cooking up these decompositions include the holdings reported periodically by funds, the weights in various benchmark indexes, the returns of the underlying securities or benchmark indexes and the reported returns of the funds. These ingredients can be combined with other data on the funds' characteristics and those of the stocks held. The first three ingredients typically abstract from any costs, while the reported returns of the funds are typically net of funds' expense ratios and trading costs. Perhaps the simplest example of this approach is the "return gap" measure of

Kacperczyk, Sialm and Zheng (2008), defined as $R_p - x(\Omega)'R$, where R_p is the return reported by the fund. The return gap captures trading within the quarter, trading costs and funds' expenses. Kacperczyk et al. find that a fund's return gap is persistent over time and has some predictive power for future fund performance.

A well-cited decomposition is proposed by DGTW (1997), where each security i held in a fund gets its own benchmark return, R_t^{bi} at each period, t, as described above. In addition, following Grinblatt and Titman (1993), the fund is assigned a set of benchmark weights equal to its actual holdings reported k periods before: $x_{i,t-k}$. The Grinblatt and Titman (1993) measure is then decomposed by adding and subtracting various bits to obtain:

$$DGTW_{t+1} = \Sigma_i \, x_{it} \, (R_{i,t+1} - R_{t+1}^{bi}) + \Sigma_i \, (x_{it} \, R_{t+1}^{bi} - x_{i,t-k} R_{t+1}^{bi(t-k)}) + \Sigma_i \, x_{i,t-k} \, R_{t+1}^{bi(t-k)} \qquad (12)$$

where $R_{t+1}^{bi(t-k)}$ is the benchmark return associated with security i at time t-k. The first term is interpreted as "selectivity," the second term as "characteristic timing" and the third as the return attributed to the style exposure. One can imagine many decompositions along these lines, and I expect to see many clever examples in subsequent research. Such decompositions come with potentially important caveats for their interpretation, however, as outlined in the next section.

4. ISSUES

This section describes some of the main issues addressed in the literature on investment performance evaluation, with a view toward future work in the area.

4.1 INTERPRETING PERFORMANCE MEASURES

4.1.1 INTERPRETING ALPHA

The fundamental literature on the normative interpretation of alpha mostly appeared in

the 1970s and 1980s and has largely died out since then. However, it did not die out because the issues were resolved. Ferson (2009) summarizes two basic questions about alpha: (1) When faced with a fund that has a positive (negative) alpha, should the investor want to buy (sell) that fund? (2) If a manager has superior information, will he or she generate a positive alpha? Ferson (2009) argues that the SDF alpha provides the best resolution of for basic questions. The first question was addressed above. On the second question, he shows that if the indirect value function of the manager is quadratic in wealth, as it would be in continuous time or under conditional normality given the manager's information, then a manager with superior information generates a positive alpha (Proposition II). However, most of the current literature on investment performance does not use SDF alphas, so it is important to understand the limitations of alpha.

As to the first question, whether an investor would wish to buy a positive-alpha fund, the simplest intuition is taught with the CAPM and proved by Dybvig and Ross (1985b, Theorem 5). At the margin, buying a positive-alpha fund to combine with a mean-variance inefficient market portfolio and cash can "beat the market" in a mean variance sense (higher mean return given the variance). However, this does not extend to a discrete investment change. Jobson and Korkie (1982) show that given an inefficient index, a portfolio with weights proportional to the vector of assets' alphas, premultiplied by an inverse covariance matrix (the optimal orthogonal portfolio), can be combined with the index to generate a mean variance efficient portfolio. The weight in the optimal orthogonal portfolio for a positive alpha asset can be negative (Gibbons Ross and Shanken (1989), Table VII). So, even if a positive alpha is attractive at the margin to a mean-variance investor, it might not imply buying a positive alpha fund given a realistic discrete response.

The early literature on investment performance was influenced by the CAPM and focused on the mean variance efficiency of the benchmarks. Chen and Knez (1996) and

Dalhquist and Soderlind (1999) used mean variance efficient benchmarks and SDF alphas, but did not assume the CAPM. This raises the question: When does a mean-variance efficient portfolio provide an Appropriate Benchmark?

A portfolio R_B is mean variance efficient if and only if it maximizes the correlation to the SDF (e.g., Ferson, 1995). That is, we can write the SDF as: $m = a + b R_B + u$, with $E(uR|Z)=0$ (the coefficients a and b may be functions of Z). Thus, $Cov(Rp,m|Z) = bCov(Rp,R_B|Z) + Cov(Rp,u|Z)$ and $Cov(R_B,m|Z)=bVar(R_B|Z)$. Substituting, we see that the covariance of Rp with m is equal to the covariance of a conditional beta-weighted combination of R_B and a risk-free asset with m, only if $Cov(Rp,u|Z)=0$. This occurs if either the SDF is exactly linear in R_B (u=0) or if R_B is conditionally mean variance efficient in the more inclusive set of assets (R, Rp). The latter condition implies that the managed fund would have a zero alpha, so the performance measurement would be trivial. Thus, we can justify a mean variance efficient benchmark only in the CAPM. Outside of this setting a mean variance efficient portfolio is not an Appropriate Benchmark.

The second question is whether a manager with superior information will produce a positive alpha. Mayers and Rice (1979) argued for an affirmative answer, assuming complete markets, quadratic utility and the CAPM. But Dybvig and Ingersoll (1982) showed that you can't marry complete markets with quadratic utility because it leads to negative state prices, and Verrechia (1980) gave a counterexample to the more general proposition that the informed earn higher returns than the uninformed expect, based on the Mayers and Rice set up. Dybvig and Ross (1985b) generalized the Mayers and Rice result to avoid the complete markets assumption but assumed that the manager has no information about the mean or variance of the uninformed client's portfolio. Connor and Korajczyk (1986) obtain similar results in an asymptotic APT, on the assumption that the investor gets an independent signal about only the APT residuals of one asset.

There are many examples where an informed manager does not generate a positive alpha. Dybvig and Ross (1985a) and Grinblatt and Titman (1989) show that a manager that is a positive market timer can generate a negative alpha. Dybvig and Ross (1985a) and Hansen and Richard (1987) show that a portfolio can be mean variance efficient given the informed manager's knowledge, but appear mean variance inefficient to the uninformed client. Thus, generally measures of alpha should be interpreted with caution. The SDF alpha seems to be on the most solid theoretical footing, and should probably get more attention than it has in the literature.

4.1.2 INTERPRETNG WEIGHT-BASED MEASURES

The two fundamental questions about alpha discussed in the last subsection apply to weight-based performance measures as well. The first question is whether a positive weight-based measure means an investor would want to buy the fund. We argued that a positive SDF alpha provides the best affirmative answer. Equation (3) shows that the SDF alpha is the sum of the covariances of the manager's weights with the future "abnormal" returns of the assets, mR. This is not what the literature on weight-based performance measures has looked at. It would be interesting to see results from using Equation (3) directly. The results are likely to differ from the existing test even if the SDF is assumed to be linear in factors -- which leads to beta pricing models (e.g., Ferson, 1995) -- because then the covariances of portfolio weights with the products of returns and factors appear in Equation (3).

With additional assumptions we can go further. Let R_B denote an appropriate benchmark. Expanding the expectation of the product in $\alpha_p \equiv E(m\,[R_p - R_B]|Z)$ we have:

$$\alpha_p = E(m|Z)\,E(R_p - R_B|Z) + Cov(m;\,[R_p - R_B]|Z) \tag{13}$$

$$= E(m|Z)\,\{\,E(x(\Omega)-x_B|Z)'\,E(R|Z) + Cov([x(\Omega)-x_B]'R|Z)\},$$

Where the second line uses $Cov(m; [R_p - R_B]|Z)=0$ for an Appropriate Benchmark. If we further assume that $E(x(\Omega)|Z) = x_B$, then the first term in {.} above is zero and we see that the remaining term is the weight-based measure of Equation (10), and is proportional to the SDF alpha. Note that if an Appropriate Benchmark satisfying $x_B = E(x(\Omega)|Z)$ is not used then the first term will not equal zero and the weight-based measure will not be proportional to the SDF alpha. This provides a strong, and I think new theoretical motivation for using benchmarks in weight-based performance measures. Note that lagged weights are not good benchmarks unless the expected value of the future weight is the lagged weight.

The second question, whether an informed manager will generate a positive weight-based measure, was addressed in equation (9) by Grinblatt and Titman (1993), who show that the condition applies if all of the holdings in a fund are included in the calculation. For example, a manager with information may overweight some assets and underweight others for hedging purposes. There may be investment ability, while the covariances on subsets of stocks are negative. It is generally not justified to apply the measures to subsets of stocks, although this is done with increasing frequency in the literature. This problem is analogous to mismeasurement of the market portfolio in the CAPM.

4.1.3 MARKET TIMING-ADJUSTED SELECTIVITY

Early studies attempt to distinguish security selection versus market timing abilities on the part of fund managers. For example, in the Merton Henriksson market timing model of Equation (8), the intercept has been naively interpreted as a measure of "timing-adjusted" selectivity performance. This only makes sense if the manager has "perfect" market timing, defined as the ability to obtain the option-like payoff at zero cost. In reality no one has perfect timing ability, and the interpretation of a_p as timing adjusted selectivity breaks down. For

example, a manager with some timing ability who picks bad stocks may be hard to distinguish from a manager with no ability who buys options at the market price. Indeed, without an estimate of the market price of a put option on the market index, the intercept has no clean interpretation. However, if the prices of options are used in the analysis, the intercept can be modified to obtain a clean measure of total performance, as shown by Aragon and Ferson (2006).

Similar issues arise in interpreting the Treynor-Mazuy market timing model of Equation (7). The intercept in the Treynor-Mazuy model does not capture the return in excess of a benchmark portfolio because r_m^2 is not a portfolio return. However, this model can also be modified to capture the difference between the return of the fund and that of an Appropriate Benchmark, as shown by Aragon and Ferson (2006). These refinements of the models have yet to be tested empirically.

4.2 MARKET EFFICIENCY

Investment performance is closely related to the issue of the informational efficiency of markets, as summarized by Fama (1970). I offer an updated interpretation of efficiency using the SDF approach. As emphasized by Fama (1970), any analysis of market efficiency involves a "joint hypothesis." There must be an hypothesis about the equilibrium model and also an hypothesis about the informational efficiency of the markets. These can be described using Equation (1). The model of market equilibrium amounts to a specification for the stochastic discount factor, m_{t+1}. For example, the Capital Asset Pricing Model of Sharpe (1964) implies that m_{t+1} is a linear function of the market portfolio return (e.g., Dybvig and Ingersoll, 1982), while multibeta asset pricing models imply that m_{t+1} is a linear function of the multiple risk factors.

Fama describes increasingly fine information sets in connection with market efficiency. Weak-form efficiency uses the information in past stock prices to form portfolios of the assets. Semi-strong form efficiency uses variables that are obviously publicly available, and strong form

uses anything else. The different information sets described by Fama (1970) amount to different assumptions about what information is contained in Z_t and what is therefore legitimately used as a lagged instrument. Weak-form efficiency says that past stock prices can't be used to generate alpha while semi-strong form efficiency says that other publicly available variables won't generate alpha. In performance evaluation, if we find that a manager has a positive alpha in a conditional model that controls for public information, this rejects a version of the joint hypothesis with semi-strong form efficiency. If investors can use the past returns of fund to earn risk-adjusted abnormal returns, this rejects a version with weak-form efficiency. If we don't question the model for m_{t+1} (and the associated Appropriate Benchmark) then we may interpret such evidence as a rejection of the informational efficiency part of the joint hypothesis.

In summary, informational efficiency says that you can't get an alpha different from zero using any information in Z_t. Since alpha depends on the model through m_{t+1}, there is always a joint hypothesis at play. Any evidence in the literature on market efficiency can be described in terms of the joint hypothesis; that is, the choice of m_{t+1} and the choice of the information in Z_t.

4.3 FUND MANAGERS' INCENTIVES AND INVESTOR BEHAVIOR

The incentives of fund managers to act on behalf of their investor-clients has long been a central topic for both theory and empirical research. Theoretically, delegated portfolio management is a subset of principle-agent models, where risk-averse investors are the principals and portfolio managers are the agents (e.g. Bhattacharya and Pfeiderer, 1985). The agents may generate a private information signal about future asset returns, and can control the scale of their unobserved action in response to the signal; for example affecting both the mean and the variance of the managed portfolio return. Portfolio managers have career concerns, and there may be multiple layers of agency, for example in pension funds and mutual fund families.

Asymmetric information lies at the heart of delegated portfolio management problems,

and it complicates things by its very presence. For example, if a manager with information knows that he will be evaluated based on the unconditional mean and variance of the portfolio return, he may be induced to form a portfolio that uses his information to maximize the unconditional mean, relative to the unconditional variance. Ferson and Siegel (2001) derive such unconditionally efficient strategies and discuss their properties as a function of the manager's information.   A manager with superior information could maximize essentially the same objective as the investor/client, and yet be seen as delivering an inferior return from the client's perspective (Dybvig and Ross, 1985a).

The incentives of portfolio managers are induced by the compensation schemes they face. Mutual fund management companies are restricted by the Investment Companies Act of 1940 from contracts with their mutual funds that include asymmetric compensation, where management firms earn more on the upside than they lose on the downside. Such schemes may induce managers to take on more risk than their investors would prefer (e.g. Starks (1987), Grinblatt and Titman, 1989c). Hedge funds face no such restrictions. The compensation of a typical hedge fund manager is a fixed percentage of the assets, say 2%, plus a share of the fund's returns, often after the fund performance exceeds a "high water mark."   Even if the management company for a mutual fund is not paid such an incentive fee, some individual managers can be and are paid this way (Elton, Gruber and Blake, 2003). Panageas and Westerfield (2009) consider the implications of a long horizon on the part of a fund manager with asymmetric incentives.

Performance benchmarks are common in the fund management industry. A benchmark may isolate the compensation of the manager from shocks captured by the benchmark that are beyond his control, and may help to separate investment skill from luck, because the variance of the difference between the fund return and that of the benchmark is typically much smaller than the variance of the portfolio return. Admati and Pfleiderer (1997) point out that it is difficult to

align the incentives of managers and investors by contracting on the differential performance. Still, explicit benchmarks are common in the pension fund industry, and since 1999 are reported by mutual funds. Roll (1992) derives the optimal response of a manager with a mean-variance utility function defined over the portfolio return net of a benchmark and Becker et al. (1999) study market timing in this setting. Their estimates suggest that asset allocation style mutual funds behave as highly risk averse, benchmark oriented investors. Chiang (2009) extends Roll's analysis for explicit conditioning information and Brennan (1993) and Gomez and Zapatero (2003) study the equilibrium implications of benchmark investors.

The incentives of fund managers are linked to the aggregate behavior of fund investors through the levels of assets under management. The flows of new money can therefore induce incentives for fund managers. Ippollito (1989) observed that mutual funds whose past returns were relatively high tended to attract relatively more new money over the next year. Evidence that investor flows chase recent high returns is found by Sirri and Tufano (1998), Chevalier and Ellison (1997) and others. Del Guercio and Tkac (2002) find that mutual fund investors pay more attention to simple measures of relative return than to more complex measures like alpha, in directing their new money flows.

Studies of the relation between flows and performance suggest an interesting nonlinear shape. Funds with the highest returns on average realize the largest subsequent inflows of new money, while funds with performance below the average do not experience withdrawals of a similar magnitude. Nonlinearity in the flow-performance relation creates an incentive for funds akin to that of a call option, even if the manager's compensation is a fixed fraction of the assets under management.

The empirical strength of the nonlinear flow-performance relation is relatively weak, considering the attention it has received in the literature. The flow-performance relation is not found to be convex for pension funds (Del Guercio and Tkac, 2002) or private equity funds

(Kaplan and Schoar, 2005). Kim (2009) recently finds that for mutual funds, the convexity that appeared in earlier studies may have largely disappeared after about 2000. Research in this area is likely to be more persistent.

Another stream of research focuses on new money flows as a window into the rationality of fund investors' behavior. Gruber (1996) forms portfolios of mutual funds, weighted according to their recent new money flows. He finds that the new money earns higher average returns than the old money invested in equity style funds. This "smart money" effect is confirmed by Zheng (1999). However, Frazzini and Lamont (2008) find that over longer holding periods the performance associated with new money flows is poor. Berk and Green (2004) present a model in which rational investors' flows chase past fund performance and in equilibrium earn no future abnormal returns. As new data on individual investors' investment behavior becomes available in the future, I expect to see much more work on how investors choose their mutual fund investments.

4.4 INTERIM TRADING

Interim trading refers to a situation where fund managers trade within the period over which returns on the fund are measured. With monthly fund return data interim trading definitely occurs, as money flows into funds and trades occur typically each day. The problems posed by interim trading were perhaps first studied by Fama (1972), and later by Goetzmann, Ingersoll and Ivkovic (2000), Ferson and Khang (2002) and Ferson, Henry and Kisgen (2006), who argue that if derivatives prices can be replicated by high-frequency trading, then the interim trading problem captures the problems posed by funds' explicit derivatives holdings. Interim trading also encompasses "risk shifting" behavior, as studied by Brown, Harlow and Starks (1996) and Koski and Pontiff (1999), as well as the manipulation of performance measures as studied by Goetzmann, Ingersoll, Spiegel and Welch (2007). Here, investors are assumed to

only pay attention to returns over a calendar year. Investment managers, trading more frequently, can then respond within the year to the returns they generate earlier in the year and to public information during the year, with the goal of influencing their annual performance numbers.

Consider an example where returns are measured over two periods, but the manager of a fund trades each period. The manager has neutral performance, but the portfolio weights for the second period can react to public information at the middle date. If only two-period returns can be measured and evaluated, the manager's strategy may appear to partially anticipate events at the middle date. For example, if market volatility falls at the middle date and the fund reacts by holding more cash, the fund's two-period return will look like that of a somewhat successful volatility timer.

Ferson and Khang (2002) conduct some experiments to assess the extent of interim trading bias. Even though their hypothetical portfolios trade only two times between each quarterly observation date, they find that the interim trading bias has a huge impact on returns-based performance measures. The average Jensen's alpha of the strategies, computed relative to the value-weighted CRSP market, is 1.03% per quarter with an average t-ratio of 3.64. Thus, by mechanically trading with public information, it is possible to generate large and economically significant alphas.

Goetzmann, Ingersoll and Ivkovic (2000) address interim trading bias by simulating the returns generated by the option to trade between return observation dates. Ferson, Henry and Kisgen (2006) use continuous-time models to address the problem. They show that if the right time-aggregated SDF is used in an SDF alpha, the interim trading bias is avoided. A time-additive utility function time aggregates in such a way as to avoid the bias, which is another strong motivation for using SDF alphas. Goetzmann et al. (2007) describe special cases.

A weight-based measure avoids interim trading bias by examining the weights at the

beginning of the period, ignoring the interim trading. Of course, managers may engage in interim trading based on superior information to enhance performance, and a weight-based measure will not record these interim trading effects. Thus, the cost of avoiding bias is a potential loss of power. Ferson and Khang (2002) evaluate these tradeoffs, and conclude that the conditional weight-based measure is attractive.

Brown, Harlow and Starks (1996) find that those funds that are performing relatively poorly near the middle of the year seem to increase their risk in the last five months of the year, as if to maximize the value of the option-like payoff caused by fund flows. Funds whose performance is relatively high near the middle of the year seem to lower their risk, as if to "lock in" their position in the performance tournament.

Koski and Pontiff (1999) examine the use of derivatives by mutual funds and find evidence of risk-shifting similar to Brown, Harlow and Starks (1996), but little evidence that the risk shifting is related to the use of derivatives. Busse (2001) argues that the evidence in the earlier studies is exaggerated by a bias, related to return autocorrelations, in estimating the standard errors of monthly returns. Using daily data he finds no evidence for risk shifting behavior. Goriaev, Nijman and Werker (2005) also find that the evidence for risk shifting is not robust, and Kim (2009) finds less evidence for risk-shifting when the convexity in the flow-performance relation is weaker, such as after 2000.

4.5 SKILL VERSUS LUCK

The task of distinguishing between skilled fund managers and lucky ones is, of course, the core of investment performance evaluation. Classical statistical analysis addresses the question of whether an estimated performance measure like alpha is significantly different from zero. However, inference is complicated by several features of managed fund data. There are a large number of funds and often a short time-series of returns, which makes time-series

asymptotics often unreliable. Many funds enter and leave the standard data bases during any sample period, raising issues of survivorship and selection biases. Some funds leave relatively short time series. Fund returns tend to be nonnormal and funds are heterogeneous in their volatility, autocorrelation and skewness.

Analyses often focus on the ex post extremes of measured performance, asking for example if the "best" managers in a sample are skilled. This question implies a multiple comparisons analysis. For example, we expect the top five out of 100 independent managers with no skill to deliver a "significant" alpha (at the 5% level). Simple approaches to adjust for the number of funds examined based on normality, such as the Bonferroni p-value, do not account for the correlation across funds, nonnormality or heterogeneity.

Recently studies have applied bootstrap resampling methods to the cross-section of funds in order to address these statistical issues. Simulating the cross-sectional distribution of performance statistics, it is possible to ask whether the estimated alphas (for example) of the top 10% of the funds in a sample are significantly larger than what is expected when the "true" alphas are zero. Kowsowski, Timmermann, Wermers and White (2006) conclude that the top 10% of active, US growth-style mutual fund alphas reflect skill. Cuthbertson, Nitzshe and Sullivan (2008) find similar results for UK equity unit trusts. Fama and French (2009) argue that these results are biased by failing to capture uncertainty in the factors. Bootstrap the factors along with the funds' residuals, they find little evidence to reject the null that the extreme alphas are due to luck. Barras, Scaillet and Wermers (2009) extend the bootstrap approach from the null that all the "true" alphas are zero to the null that only a fraction are zero. They estimate the fraction to be about 75%, while almost one fourth of the funds have truly negative alphas. I expect to see more work along these lines in the future.

Recent research exploits the availability of data on fund characteristics and holdings to improve the power of tests. For example studies such as Chen, Jegadeesh and Wermers (2000),

Cohen, Coval and Pastor (2005), Kacperczyk, Sialm and Zheng (2005, 2008), Gaspar, Massa and Matos (2006), Cremers and Petajisto (2009), Kacperczyk and Seru (2007), Cohen, Frazzini and Malloy (2008) and Huang, Sialm and Zhang (2008) combine data to find evidence that some types of funds have stock selection skills at some times. These studies typically rely on classical statistical inferences. I think that it should prove useful to extend the bootstrap analysis of statistical significance to the context of these studies.

Another response to the low statistical power of classical performance measurement is a Bayesian approach. Bayesian analyses of funds are provided by Baks, Metrick and Wachter (2000), Jones and Shanken (2005), Avramov and Wermers (2006) and others. Pastor and Stambaugh (2002) for example, exploit the correlations of measured fund returns with those of "seemingly unrelated" assets to refine portfolio decisions involving mutual fund returns.

4.6 LIQUIDITY

There are several aspects of liquidity that are relevant to the problem of measuring investment performance. First, mutual funds provide liquidity to their investors in the form of an option to trade at the closing net asset value per share each day. This liquidity option was discussed above in the context of market timing and the late trading scandals. Second, the liquidity of the assets held by funds can be important in several respects. Funds may earn liquidity premiums for holding illiquid assets, and these may be partially distributed to fund investors (e.g. Aragon, 2007). Funds may be forced to sell illiquid assets if they experience large outflows, which may affect asset prices (Coval and Stafford, 2007) and lead to systemic effects similar to bank runs, as experienced during the financial crisis of 2008.

A third aspect of illiquid assets is that they are likely to trade asynchronously. Stale prices in reported net asset values create statistical biases in measured fund returns. For example, funds' betas and volatility are likely to be biased downward, so Sharpe ratios and

alphas may be biased upward (Asness, Krail and Liew (2001), Getmansky, Lo and Makarov, 2004). If the staleness varies over time in relation to market factors, funds' market timing coefficients may be biased (Chen et al., 2009). See Qian (2009a) for an analysis of biases due to stale prices in equity mutual funds. I suspect that all of these liquidity-related issues present further opportunities for further interesting research.

4.7 BEYOND EQUITY FUNDS

While most of the early research on investment performance concentrated on equity funds, there are now developed and developing bodies of research on hedge funds, bond funds, pension funds and a range of other investment vehicles. Consider bond funds. In 2003 the total net assets of U.S. bond funds exceeded 1.2 trillion dollars, about 1/6 the amount in equity-style mutual funds and similar to the value of hedge funds. Large amounts of fixed-income fund assets are also held by pension funds, trusts and insurance company accounts. However, the number of academic studies of bond funds is relatively small. Elton, Gruber and Blake (1993) and Elton, Gruber and Blake (1995) were the seminal academic studies of the performance of bond style mutual funds. Comer, Boney and Kelly (2005) study timing ability in a sample of 84 high quality corporate bond funds, 1994-2003, using variations on Sharpe's (1992) style model and Comer (2006) measures bond fund sector timing using portfolio weights. Ferson, Henry and Kisgen (2006) bring modern term structure models to the problem of measuring bond fund performance and Moneta (2009) applies weight-based performance measures to a large sample of US bond funds. Chen, Ferson and Peters (2009) study the ability of US bond funds to time several factors related to bond markets.

The recent growth in hedge fund assets and the significant attention devoted to hedge funds in the popular press has increased the interest in hedge fund performance. Like mutual funds, hedge funds are open-ended investment companies that pool dollars from a group of

investors. Unlike mutual funds, hedge funds are exempt from the Investment Company Act. Hedge fund performance presents several interesting issues, and there is now a large body of evidence on hedge fund performance, styles and strategies. The performance of hedge funds looked promising when academics studies first began to explore it empirically, as hedge funds delivered large alphas in traditional linear beta models. However, as this literature has matured, it seems that the large alphas of hedge funds can be explained through a combination of data biases, such as survivor selection and backfilling, dynamic trading and nonlinear payoffs, asset illiquidity and smoothed reported prices. A more detailed review of the literature on hedge funds is provided by Aragon and Ferson (2006).

There is a modest body of work on Pension funds. Coggin, Fabozzi and Rahman (1993) study the market timing and stock picking ability of US equity style pension funds and provide references to the few academic studies that were available at the time. Lakonishok, Shleifer and Vishny (1992) find that the typical pension fund earned average returns about 1% below that of the SP500 index over 1983-1989, which they interpret as evidence of poor performance. However, Christopherson, Ferson and Glassman (1998b) show that pension funds tend to hold smaller stocks than those of the SP500 on average. They also find that Lakonishok, Shleifer and Vishny evaluated their funds during a sample period when small stocks returned less than large stocks. Using a more appropriate style-based benchmark that controls for the market capitalization of the stocks, Christopherson et al. found evidence for positive pension fund alphas. Ferson and Khang (2002) find that returns-based alphas of their sample of pension funds are positive, but smaller than the potential effects of interim trading bias. Their conditional weight-based measures find that the pension funds have neutral performance. If the importance of defined benefit pension plans continues to decreases in the future, I do not see this as a growth area for future research.

The use of portfolio holdings data in combination with other data on funds and the stocks

that they hold is a burgeoning area of research that I expect will continue to grow rapidly. Studies combining these data have broadened the range of issues in the purview of the investment performance literature. Examples of these issues include the effects of regulatory requirements such as disclosure (e.g. Wermers, 2001), predatory strategic behavior (Chen, Hansen, Hong and Stein, 2008), cross-subsidization in multi-product firms (e.g. Gaspare, Massa and Matos, (2006), Reuter, 2006), liquidity and trading costs (Christophersen, Keim and Musto, 2007), corporate governance issues such as the role of boards of directors (e.g. Qian, 2009), the effects of institutional shareholdings on corporate merger decisions (Gaspare, Massa and Matos, 2005), operating performance (Ferreira and Matos, 2008) and other corporate decisions, the roles of business and personal networks (Cohen, Frazzini and Malloy, 2008), the effects of industrial organization such as customer-supplier relations (Huang and Kale, 2009) and much more.

5. FUTURE WORK

This review has made several suggestions for future work. In general, measures of alpha should be interpreted with caution. The SDF alpha seems to be on the most solid theoretical footing, and should probably get more attention than it has in the literature. Market timing models can be refined to capture the difference between the return of the fund and that of an appropriate benchmark portfolio, as shown in Aragon and Ferson (2006). The various aspects of liquidity deserve more research, and more work on the incentives and compensation of portfolio managers seems to be in the cards.

I suggest several refinements for portfolio holdings-based performance measures. Equation (3) shows that the SDF alpha is the sum of the covariances of the manager's weights with the future "abnormal" returns of the assets, mR. It would be interesting to see this measure estimated. The literature has not yet developed methods to separate risk aversion from signal quality in weight-based measures. We showed that if an Appropriate Benchmark satisfying $x_B = E(x(\Omega)|Z)$ is not used then weight-based measures will not be proportional to the SDF alpha.

This provides a strong, and I think new theoretical motivation for using benchmarks in weight-based performance measures. Benchmarks may also improve the statistical properties of estimated measures, although the relevant statistical issues have yet to be worked out.

Recent research combines data on fund characteristics, holdings and returns to improve the power of tests and to broaden the set of issues that can be studied. However, work in this area typically relies on asymptotic statistical inference. I think that it should prove useful to extend the bootstrap analysis of the cross section, as has been recently applied to alphas, to the context of these studies. Similarly, few studies have used simulation to address the significance of timing ability for subsets of managers.

5.1. A MORE BALANCED TREATMENT OF COSTS

A manager may be able to generate higher returns than an Appropriate Benchmark before costs and fees, yet after costs investors' returns may be below the benchmark. If a fund can beat the benchmark on an after cost basis, Aragon and Ferson (2006) say that the fund adds value for investors, to distinguish this situation from one where the manager has investment ability, but either extracts the rents to this ability in the form of fees and expenses, or dissipates it through trading costs. It would be useful to have cleaner measures of investment performance, both on a before-cost and an after-cost basis.

Mutual fund returns are measured net of all the expenses summarized in the funds' expense ratio and also the trading costs incurred by the fund. Funds' trading costs can be substantial, both for managed funds and even for common "passive" benchmarks. Sometimes, transactions fees are paid into the assets of the fund by new investors to compensate existing shareholders for the costs of buying and selling the underlying assets. Funds sometimes also charge additional "load fees," such as those paid to selling brokers -- measured fund returns do not account for these additional charges.

Most performance measures are crude in their treatment of investment costs and fees. In most academic studies the benchmark strategy does not reflect its costs. For example, the CRSP indexes pay no costs when their composition changes. Thus, the typical performance measure compares apples to oranges. I think it makes sense to modify current measures to reflect the costs of trading the benchmarks. Then, we would have a better sense of the performance after costs.

Weight based performance measures are fairly clean before cost measures, as they amount to the average difference between a hypothetical portfolio and a benchmark that pays no costs. The hypothetical portfolio is constructed from a snapshot of the fund's weights and abstracts from expenses and trading costs.

Measuring the managed portfolio's returns and the performance benchmark returns on a cost-equivalent basis can get complicated. A big problem is that the incidence of many costs is likely to be different for different investors. For example, a pension plan pays no income taxes on the dividends or capital gains generated by a portfolio, so the manager and the plan client may care little about the form in which the gains are earned. An individual investor may be taxed more favorably on capital gains than on dividends, and the relative tax cost may depend on the investor's income profile. This implies that different investors may view the performance of the same fund in different ways.

5.2 CLIENTELE-SPECIFIC PERFORMANCE MEASURES

Fundamentally, performance measures always amount to some specification of the benchmark. It is appealing to seek benchmarks about which there can be unanimity, as the performance measure can then be relevant for all investors. However, this is more of a modeling fiction than a practical reality. While some investment managers try hard to divine their clients' preferences, the right client-specific SDF is likely to be difficult to measure in practice. It may

be practical however, to closely approximate the "right" alpha by identifying cohorts of investors based on investor characteristics, who would be expected to have similar SDF's. It would be interesting to see how different the measures are for different cohorts. The literature using firm-specific characteristics, country-specific characteristics and more detailed data about individual investors has developed dramatically over the past two decades, so maybe the time is right for investor-characteristics-based performance measures.

The theoretical ambiguities in the interpretation of alpha found in the early literature are largely resolved when alpha is defined relative to the client's preferences. Properly accounting for costs and taxes in performance comparisons is also client specific. In evaluating managed portfolios, one size does not fit all. A challenge for future research, therefore, is to identify and characterize meaningful investor clienteles and to develop performance measures specific to the clienteles.

6. CONCLUSIONS

This was a selective overview of the rapidly developing literature on investment performance evaluation. I show that the stochastic discount factor approach unifies the issues and offers some new insights. I summarized the significant forces and contributions that have brought this field of research to its current state of knowledge and made some suggestions for future research. These suggestions include refinements to portfolio holdings-based performance measures, a more balanced treatment of costs and clientele-specific measures of investment performance.

References

Admati AR, Bhattacharya S, Pfleiderer P, Ross SA. 1986. On Timing and Selectivity. *Journal of Finance* 41: 715-30

Admati AR, Pfleiderer P. 1997. Does it all add up? Benchmarks and the compensation of portfolio managers. *Journal of Business*. 70: 323-50

Aragon G. 2007. Share restrictions and asset pricing: Evidence from the hedge fund industry. *Journal of Financial Economics* 83: 33-58

Aragon G, Martin S, A unique view of hedge fund derivative useage: Safeguard or speculation? Working paper, Carnegie Mellon University.

Aragon G, Ferson W. 2006. Portfolio Performance Evaluation. *Foundations and Trends in Finance* 2:2: 1-111

Asness C. Krail R, Liew, J. 2001. Do hedge funds hedge? *Journal of Portfolio Management*. 28: 6-19

Avramov D, Wermers R. 2006. Investing in mutual funds when returns are predictable. *Journal of Financial Economics* 81: 339-77

Baks K, Metrick A, Wachter J. 2001. Should investors avoid all actively-managed funds? A study in Bayesian performance evaluation. *Journal of Financial Economics* 56(1): 45-85

Barras L, Scaillet O, Wermers R. 2009.   False discoveries in mutual fund performance: Measuring luck in estimated alphas, forthcoming, *Journal of Finance*.

Beja A. 1971. The structure of the cost of capital under uncertainty. *Review of Economic Studies*. 4:359-369

Berk J, Green RC. 2004. Mutual Fund Flows and Performance in Rational Markets. *Journal of Political Economy* 112: 1269-95

Becker C, Ferson W, Myers D, Schill M. 1999. Conditional Market timing with Benchmark investors. *Journal of Financial Economic*s. 52: 119-48

Bhattacharya S, Pfleiderer P. 1985. Delegated portfolio Management. *Journal of Economic Theory*. 36:1-25

Blake C, Elton E, Gruber M. 2003. Incentive fees and mutual funds. *Journal of Finance*. 58(2):

779-804

Brennan MJ. 1993. Agency and Asset Prices. Working paper. UCLA.

Brinson G, Hood L, Bebower G. 1986. Determinants of Portfolio Performance. *Financial Analysts Journal* 42: 39-44

Brown K, Harlow V, Starks L. 1996. Of tournaments and temptations: An Analysis of managerial incentives in the mutual fund industry. *Journal of Finance.* 51(1): 85-110

Busse J. 1999. Volatility timing in mutual funds: Evidence from daily returns. *Review of Financial Studies.* 12:1009-41

Busse J. 2001. Another look at mutual fund tournaments. *Journal of Financial and Quantiative Analysis.* 36: 53-73

Carhart, M. M., 1997. On persistence in mutual fund performance. Journal of Finance 52: 57-82.

Carlson R S. 1970. Aggregate Performance of Mutual Funds, 1948-1967. *Journal of Financial and Quantitative Analysis.* 5: 1-32

Carino D, Christopherson J, Ferson W. 2009. *Portfolio Performance Measurement and Benchmarking.* McGraw Hill Books

Chen HL, Jegadeesh N, Wermers R. 2000. An examination of the stockholdings and trades of mutual fund managers. *Journal of Financial and Quantitative Analysis* 35:43-68

Chang EC, Lewellen WG. 1984. Market timing and mutual fund investment performance. *Journal of Business.* 57: 57-72.

Chen J, Hanson S, Hong H, Stein J. 2008. Do hedge funds profit from mutual fund distress? Working paper, University of California at Davis.

Chen N, Copeland T, Mayers D. 1987. A comparison of single and multifactor performance methodologies. *Journal of Financial and Quantitative Analysis* 22: 401-17

Chen N, Roll R, Ross S. 1986. Economic Forces and the Stock Market. *Journal of Business* 59: 383-403

Chen Y, Ferson W, Peters H. 2009. Measuring the Timing Ability of Fixed Income Funds, forthcoming, *Journal of Financial Economics.*

Chen Z, Knez PJ. 1996. "Portfolio performance Measurement: theory and applications." *Review of Financial Studies.* 9: 511-56

Chevalier J, Ellison L. 1997. Risk Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy* 105: 1167-1200

Chordia T. 1996. The structure of mutual fund charges. *Journal of Financial Economics* 41: 3-39

Christopherson JA, Ferson W, Glassman DA. 1998. Conditioning manager alpha on economic information: Another look at the persistence of performance. *Review of Financial Studies*. 11: 111-42

Christophersen S, Keim D, Musto D. 2007. Valuable information and costly liquidity: Evidence from individual mutual fund trades, working paper, Wharton.

Coggin D, Fabozzi F, Rahman, S. 1993. The investment performance of U.S. equity pension fund managers. *Journal of Finance* . 48: 1039-56

Cohen RJ, Frazzini A, Malloy C. 2008. The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy* 116(5): 951-79

Cohen RJ, Coval J, Pastor L. 2005. Judging fund managers by the company they keep, *Journal of Finance* 60: 1057-96

Coles A. 1933. Can stock market forecasters forecast? *Econometrica* 1(3): 309-25

Comer G, Boney V, Kelly L. 2005. Timing the investment grade securities market: Evidence from high quality bond funds. *Journal of Empirical Finance* 16(1): 55-69

Comer G. 2006. Evaluating bond fund sector timing skill. Working paper, Georgetown University

Connor G, Korajczyk, RA. 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*. 15: 373-94

Connor G, Korajczyk, RA. 1988. Risk and return in an equilibrium APT: Applications of a new test methodology. *Journal of Financial Economics*. 21: 255-89

Copeland T, Mayers D. 1982. The value line enigma (1965-1978): a case study of performance evaluation issues. *Journal of Financial Economics* 10, 289-321

Cornell B. 1979. Asymmetric information and portfolio performance measurement. Journal of Financial Economics 7, 381-90

Coval J, Stafford E. 2007. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics* 86(2): 479-512

Cremers M, Petajisto A. 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies.* 22: 3329-65

Cutherbertson K, Nitzsche D, O'Sullivan N. 2008. UK mutual fund performance: Skill or Luck? *Journal of Empirical Finance*. 15: 613-34

Dahlquist M, Soderlind P. 1999. Evaluating portfolio performance with stochastic discount factors. *Journal of Business* 72: 347-84

Daniel K, Grinblatt M, Titman S, Wermers R. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52: 1035-58

Del Guercio D, Tkac PA. 2002. The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds. *Journal of Financial and Quantitative Analysis*. 37(4): 523-57

Dybvig PH, Ingersoll JE. 1982. Mean Variance Theory in Complete Markets. *Journal of Business* 55: 233-251

Dybvig PH, Ross SA. 1985a. Differential Information and Performance Measurement Using and a Security Market Line. *Journal of Finance*. 40: 383-99

Dybvig PH, Ross SA. 1985b. The Analytics of Performance Measurement Using and a Security Market Line. *Journal of Finance* 40: 401-16

Edelen, RM. 1999. Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics*, 53: 439-66

Elton E, Gruber MJ, Blake CR. 1993. The performance of bond mutual funds. *Journal of Business* 66: 371-403

Elton E, Gruber MJ, Blake CR. 1995, Fundamental Economic Variables, Expected returns and bond fund performance. *Journal of Finance* 50: 1229-56

Evans RB. 2003, Does alpha really matter? Evidence from fund incubation, termination and manager change, working paper, University of Pennsylvania.

Fama EF. 1970. Efficient Capital Markets: A review of theory and empirical work. *Journal of Finance* 25: 383-417

Fama EF. 1972. Components of investment performance. *Journal of Finance* 27: 551-67

Fama EF, French KR. 1996. Multifactor explanations of asset pricing anomalies. *Journal of*

Finance 51: 55-87

Fama EF, French KR. 2009. Luck versus skill in the cross section of mutual fund alpha estimates. *Journal of Finance* (forthcoming)

Farnsworth HK, Ferson W, Jackson D, Todd S. 2002. Performance Evaluation with Stochastic Discount Factors. *Journal of Business* 75: 473-504

Ferson, WE. 1995. Theory and Empirical Testing of Asset Pricing Models, Chapter 5 in Finance, Handbooks in Operations Research and Management Science, by Jarrow, Maksimovic and Ziemba (editors), Elsevier, 145-200

Ferson W. 2009. Alpha and the problem of performance measurement. Working paper, University of Southern California

Ferson W, Henry T, Kisgen D. 2006. Evaluating Government Bond Fund Performance with Stochastic Discount Factors. *Review of Financial Studies* 19(2): 423-55

Ferson W, Khang K. 2002. Conditional performance measurement using portfolio weights: Evidence for pension funds. *Journal of Financial Economics* 65: 249-82

Ferson W, Qian M. 2004, Conditional performance evaluation revisited, *Research Foundation of the CFA Institute.*1-85

Ferson W, Schadt R. 1996. Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance* 51: 425-62

Ferson W, Siegel AF. 2001. The efficient use of conditioning information in portfolios. *Journal of Finance*, 56: 967-82

Ferson W, Warther VA. 1996. Evaluating Fund Performance in a Dynamic Market. *Financial Analysts Journal* 52(6): 20-8

Ferreira J, Matos P. 2008. The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics* 88(3): 499-533

Frazzini A, Lamont O. 2008. Dumb Money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics* 88(2): 299-322

Gaspar JM, Massa M, Matos P. 2006. Favoritism in Mutual Fund Families? Evidence on strategic cross-fund subsidization. *Journal of Finance* 61(1): 73-104

Gaspar JM, Massa M, Matos P. 2005, Shareholder investment horizon and the market for corporate control. *Journal of Financial Economics 76(1): 73-104*

Getmansky M, Lo A, Makarov I., 2004. An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics* 74: 529-610

Gibbons M, Ross SA, Shanken J. 1989. A test of the efficiency of a given portfolio, *Econometrica*. 57: 1121-52

Glosten L, Jagannathan R. 1994. A contingent claims approach to performance evaluation. *Journal of Empirical Finance* 1: 133-66

Goetzmann W, Ingersoll J, Ivkovic Z. 2000. Monthly measurement of daily timers. *Journal of Financial and Quantitative Analysis*. 35: 257-90

Goetzmann W, Ingersoll J, Spiegel M, Welch I. 2007, Portfolio performance manipulation and manipulation-proof measures. *Review of Financial Studies* 29(5): 1503-46

Gomez JP, Zapatero F. 2003. Asset pricing implications of benchmarking: A two-factor model. *European Journal of Finance*. 9: 343-57

Goriaev, M., T. Nijman and B. Werker, 2005. Yet another look at mutual fund tournaments, *Journal of Empirical Finance* 12, 127-37

Griffin J, Xu J. 2009. How smart are the smart guys? A unique view from hedge fund stock holdings, *Review of Financial Studies* 22, 2531-570

Grinblatt M, Titman S. 1989a. Mutual fund performance: an analysis of quarterly portfolio holdings. *Journal of Business* 62: 393-416

Grinblatt M, Titman S. 1989b. Portfolio Performance evaluation: old issues and new insights. *Review of Financial Studies* 2: 393-422

Grinblatt M, Titman S. 1989c. Adverse risk incentives and the design of performance-based contracts. *Management Science*. 35:807-822

Grinblatt M, Titman S. 1993. Performance measurement without benchmarks: an examination of mutual fund returns. Journal of Business 60, 97-112

Grinblatt M, Titman S, Wermers R. 1995. Momentum strategies, portfolio performance and herding: A study of mutual fund behavior. American Economic Review 85, 1088-105

Gruber M. 1996. Another Puzzle: The Growth in Actively Managed Mutual Funds. *Journal of Finance* 51: 783-810

Hansen LP, Richard S. 1987. The role of conditioning information in deducing testable

restrictions implied by dynamic asset pricing models. *Econometrica* 55: 587-613

Henriksson R, Merton R. 1984. Market timing and mutual fund performance: an empirical Investigation. *Journal of Business* 57: 73-96

Huang J, Sialm C, Zhang H. 2008. Risk shifting and mutual fund performance, working paper, University of Texas

Huang L, Kale J. 2009. The effect of supplier and customer industry interrelation on mutual fund investment and performance, working paper, Georgia State University

Ippolito RA. 1989. Efficiency with costly information: A study of mutual fund performance. 1965-1984. *The Quarterly Journal of Economics* 104: 1-23

Ippolito RA. 1992. Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry. *Journal of Law and Economics* 35: 45-70

Irvine P, Nathan S, Simko P. 2004. Asset management and affiliated analysts' forecasts. *Financial Analysts Journal* (May/June): 67-78

Jagannathan R, Korajczyk R. 1986. Assessing the market timing performance of managed portfolios. Journal of Business 59: 217-36

Jensen MC. 1968. The performance of mutual funds in the period 1945 -1964. *Journal of Finance* 23: 389-46

Jensen MC. 1972. Optimal Utilization Of Market Forecasts And The Evaluation Of Investment Performance. *Mathematical Methods in Finance*, G. P. Szego, Karl Shell, eds., North-Holland Publishing Company

Jobson JD, Korkie R. 1982. Potential performance and tests of portfolio efficiency. *Journal of Financial Economics*. 10: 433-66

Jones CS, Shanken J. 2005. Mutual fund performance with learning across funds. *Journal of Financial Economics*. 78:507-52

Kacperczyk MC, Seru A. 2007. Fund manager use of public information: New evidence of managerial skills, *Journal of Finance* 62(2), 485-528

Kacperczyk MC, Sialm C, Zheng L. 2005, On the industry concentration of actively managed equity mutual funds. *Journal of Finance* 60(4), 1983-2011

Kacperczyk M.C, Sialm C, Zheng L. 2008. Unobserved actions of mutual funds, *Review of Financial Studies* 21(6), 2379-416

Kaplan S, Schoar A. Private equity performance: Returns, persistence and capital flows. *Journal of Finance*. 60(4): 1791-1823

Kim, M. 2009. Changes in mutual fund flows and managerial incentives. Working paper, University of Southern California

Kon SJ. 1983. The market timing performance of mutual fund managers. *Journal of Business* 56: 323-47

Koski J, Pontiff J. 1999. How are derivatives used? Evidence from the mutual fund industry. *Journal of Finance* 54, 791-816

Kowsowski R, Timmermann A, Wermers R, White HL. 2006. Can mutual fund "stars" really pick stocks? Evidence from a bootstrap analysis. *Journal of Finance* 61: 2251-95

Kryzanowksi L, Lalancette S, To M. 1996. Performance attribution using an APT with prespecified factors. Working paper, Concordia University

Lakonishok J, Shleifer A, Vishny R. 1992. The structure and performance of the money management industry. Brookings Papers on Economic Activity: Microeconomics. 339-91

Lehmann, BN, Modest DM. 1987. Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42: 233-66

Leland, H. 1999. Performance beyond mean-variance: Performance measurement in a nonsymmetric world. *Financial Analysts Journal* 55: 27-36

Long JB. 1974. Stock prices, inflation, and the term structure of interest rates. *Journal of Financial Economics* 1: 131-70

Mayers D, Rice EM. 1979. Measuring portfolio performance and the empirical content of asset pricing models. *Journal of Financial Economics* 7: 3-28

Merton RC. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867-87

Merton RC, Henriksson RD. 1981. On market timing and investment performance II: Statistical procedures for evaluating forecasting skills. *Journal of Business* 54: 513-34

Moneta F. 2009. Measuring bond fund performance with portfolio characteristics. Working paper, Queens University

Panageas S, Westerfield M. 2009. High water marks: High risk appetites? Convex compensation,

long horizons and portfolio choice. *Journal of Finance* 64(1): 1-36

Pastor L, Stambaugh R. 2002. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics* 63(3): 351-80

Qian, M. 2009. Stale prices and the performance evaluation of mutual funds. *Journal of Financial and Quantitative Analysis* (forthcoming)

Qian M. 2009. Is voting with feet an effective mutual fund governance mechanism? *Journal of Financial and Quantitative Analysis* (forthcoming)

Reuter JM. 2006. Are IPO allocations for sale? Evidence from Mutual funds. *Journal of Finance* 61(5): 2289-324

Roll R. 1992. A mean/variance analysis of tracking error. *Journal of Portfolio Management*. 8(4): 13-22

Roll R. 1978. Ambiguity when performance is measured by the security market line. *Journal of Finance* 33: 1051-69

Ross SA. 1976. The arbitrage pricing theory of capital asset pricing. *Journal of Economic Theory*. 13: 341-60

Sharpe WF. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*. 19: 425-42

Sharpe WF. 1992. Asset allocation: Management style and performance measurement. *Journal of Portfolio Management* 18: 7-19

Sirri ER, Tufano P. 1998. Costly Search and Mutual Fund Flows. *Journal of Finance*, 53: 1589-1622

Starks L. 1987. Performance incentive fees: an agency theoretic approach. *Journal of Financial and Quantitative Analysis* 22, 17-32

Tkac, Paula, 2008, Mutual Fund Innovation: Past, Present and Future. Working paper, Federal Reserve Bank of Atlanta

Treynor J, Mazuy K. 1966. Can Mutual funds Outguess the Market? *Harvard Business Review*, 44: 131-36

Verrecchia RE. 1980. Consensus Beliefs, Information Acquisition, and Market Information Efficiency. *The American Economic Review* 70: 874-84

Wermers R. 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs and expenses, *Journal of Finance* 55: 1655-95

Wermers R. 2001. The potential effects of more frequent portfolio disclosure on mutual fund performance. *Investment Company Institute Perspective* 7, 1-12

Zheng L. 1999. Is money smart?: a study of mutual fund investors' fund selection ability. *Journal of Finance* 54, 901-33

Zitzewitz, Eric, 2006, How widespread was late trading in mutual funds? *American Economic Review 96(2): 284-89*